HEALTH ECONOMICS RESEARCH, INC.

300 Fifth Avenue, 6th Floor
Waltham, MA 02154
(617) 487-0200
(617) 487-0202 Fax

UNIQUE PHYSICIAN IDENTIFICATION NUMBER (UPIN) VALIDATION STUDIES

Documentation For The UPIN Integrity And Multiple-UPIN Files

Submitted by:

Killard W. Adamache, Ph.D.
William R. Scott, III

Prepared for:

Ann Meadow, Sc.D.
HCFA Project Officer

October 5, 1994

This research was funded under Delivery Order No. 5 of the HCFA Master Contract No. 500-92-0020. The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of HCFA. The contractor assumes responsibility for the accuracy and completeness of information contained in this report.

TABLE OF CONTENTS

		PAGE
1.0	**INTRODUCTION**	-1-
2.0	**REGISTRY RECORD AND HCFA RECORD LAYOUTS**	-2-
3.0	**DEVELOPMENT OF THE INTEGRITY AND MULTIPLE-UPIN FILES**	-4-
	3.1 Integrity File Flags	-4-
	3.1.1 Individual Data Elements	-4-
	3.1.2 Logical Consistency	-5-
	3.1.3 Consistency of Header and Practice Setting Values	-6-
	3.1.4 Consistency Across Practice Settings	-7-
	3.2 Development of the Integrity File	-7-
	3.3 Development of the Multiple-UPIN File	-8-
4.0	**MAINTENANCE OF PROGRAMS USED TO DEVELOP THE INTEGRITY FILE**	-9-
5.0	**REPORTS CREATED FROM THE INTEGRITY AND MULTIPLE-UPIN FILES**	-10-
	5.1 Integrity File Reports	-10-
	5.1.1 Practice Setting Reports	-11-
	5.1.2 Header Reports	-12-
	5.2 Multiple-UPIN Report	-12-

TABLE OF EXHIBITS

		PAGE
EXHIBIT 1	STRUCTURE OF A RECORD FOR AN INDIVIDUAL PROVIDER WITH MULTIPLE PRACTICE SETTINGS	-14-
EXHIBIT 2	FORMATS OF REGISTRY DATA ELEMENTS	-15-
EXHIBIT 3	FLOW OF PROGRAMS TO CREATE THE INTEGRITY AND MULTIPLE-UPIN FILES	-16-
EXHIBIT 4	VARIABLE NAMES FOR REGISTRY DATA ELEMENTS	-17-
EXHIBIT 5	DEFINITION OF DATA ELEMENT INTEGRITY MEASURES	-19-
EXHIBIT 6	COMPONENTS OF KEY VARIABLES FOR DETECTING MULTIPLE-UPIN CASES	-22-
EXHIBIT 7	DATA ELEMENTS REQUIRING UPDATES OF BOUNDARY VALUES FOR FLAG VALUES: QUARTERLY FILE AND UPDATE FREQUENCY	-23-
EXHIBIT 8	INTEGRITY AND MULTIPLE-UPIN REPORTS	-24-
EXHIBIT 9	SAMPLE PRACTICE SETTING REPORT (EXPANDED)	-25-
EXHIBIT 10	SAMPLE HEADER REPORT (EXPANDED)	-26-
EXHIBIT 11	SAMPLE MULTIPLE-UPIN REPORT	-27-

1.0 INTRODUCTION

The *Physician Registry* was established in 1989 by HCFA for the purpose of issuing unique identification numbers (UPINs) to the following Medicare providers: doctors of medicine, osteopathy, dental medicine, dental surgery, podiatry and optometry, and chiropractors. To assist in the identification of individual providers, the *Registry* collects a variety of information.

Two special files containing *Physician Registry* data were created to assist HCFA, *Registry* personnel, carrier personnel and researchers in using the data. One file – *UPIN Integrity* file – contains unedited, original *Registry* data and special variables (i.e., "flags") that indicate data quality for each provider practice setting. The other file – *Multiple-UPIN* file – contains records of providers that may have more than one UPIN. These files were developed by Health Economics Research, Inc. (HER) based on its analyses of Registry data.[1]

The flag variables can be used by operations personnel to quickly obtain frequencies of the different types of non-feasible values given a data element. Researchers can similarly use the flags to obtain a quick assessment of data quality without having to set their own criteria for doing so.

The next section describes *Registry* file and record formats. It is followed by a description of the development of the flag variables and the *Integrity* and *Multiple-UPIN* files. The next section describes the changes that periodically need to be made to the programs that created the *Integrity* file. The last section describes reports that were created to assist the *Registry* and carriers in maintaining *Registry* data.

[1]Unique Physician Identification Number (UPIN) Validation Studies: Carrier Analysis, HCFA Master Contract No. 500-92-0050, Delivery Order No. 5, August 23, 1994.

2.0 REGISTRY RECORD AND HCFA RECORD LAYOUTS

In this section, the files used as basic input files in creating the *Integrity* and *Multiple-UPIN* files are described. Each quarter, Transamerica Occidental Life Insurance Company (TOLIC), the *Registry* contractor, sends an updated copy of the *Registry* file to HCFA. For each provider there is one record. A provider's record in the *Registry's* file consists of a *header* section, containing data that should be invariant across practice settings and a *practice setting* section. Within the practice setting section, there are as many practice setting segments as the provider has practice settings, both active and inactive. Because the number of practice setting segments varies by provider, a file with this type of record structure is known as a *variable-length record file*. This structure is illustrated in the top half of Exhibit 1.

There are several data elements in TOLIC's file that have a different format from the records submitted by carriers to the *Registry* (Exhibit 2). In particular, several of the date fields in the TOLIC file have a *packed* numeric format and the components of the date variables are transposed. Also included in Exhibit 2 is the value carriers were instructed to use to indicate "missing" information.

Upon receipt of the file, HCFA's Bureau of Data Management and Strategy (BDMS) creates a 515-byte flat file with fixed-length records (first item in Panel A in Exhibit 3). For each provider, BDMS creates as many records as there are practice settings in the provider's TOLIC record. Each provider record consists of a header section and a practice setting section as illustrated in the bottom half of Exhibit 1.

Selected variables created by BDMS from the TOLIC file are listed in Exhibit 4. Only those variables in the BDMS file that are used in the creation of the *Integrity* and *Multiple-UPIN* files are listed. The first column in Exhibit 4 contains the names of data elements submitted to TOLIC on carrier records. The second column contains an expanded definition of carrier data elements. The third column contains a code that indicates whether the variable is a header

variable (code=H), or whether it is a practice setting variable (code=S) on the BDMS file. The fourth column is the "field name" used in HCFA's documentation of the flat file.

From the flat file created by BDMS, HER has created an SAS dataset that has the same structure as the BDMS file. For its developmental work, HER used the April 1994 flat file. The SAS variable names assigned by HER are listed in the fifth column of Exhibit 4. The program used by HER to create the SAS dataset is available from HCFA (item 2 in Panel A of Exhibit 3). In the SAS file, dates are stored as numbers rather than in a SAS date format. All but one of the SAS programs described in Sections 3.2 and 3.3 uses this SAS dataset as its input file. (The exception is the OUTSUPER program which is the last step in creating the *Integrity* file – see Panel B of Exhibit 3.)

There are several data elements that are common to both the header and practice setting segments of the provider record. For instance, the SAS name for the header version of the date of birth is DOB while the name for the practice setting version is SETDOB. Note, however, that the values of the two versions could differ, because they are derived from their respective sections in the TOLIC provider record. TOLIC does not force the two values to equal each other. The following variables have both header and practice setting versions:

- provider name,
- date of birth,
- school code,
- graduation year,
- sanction information:
    - sanction code,
    - sanction date, and
    - sanction length.

The practice setting versions of these variables are the same as those used in the carrier records transmitted to TOLIC.

3.0 DEVELOPMENT OF THE INTEGRITY AND MULTIPLE-UPIN FILES

3.1 Integrity File Flags

The use of the *Integrity* file is centered on the flag variables created during its construction. There are two types of flag variables. The flag variables for individual variables indicate whether the value of a variable is feasible or non-feasible and, if non-feasible, its nature (e.g., retired code present). The second type indicates whether the coding of two variables being compared is consistent. The basic categories for each type of flag variable are described in the four sections that follow.

3.1.1 Individual Data Elements

Flag variables for individual variables indicate whether the value of a variable is feasible or non-feasible, and, if non-feasible, its nature (e.g., missing). For example, a date of birth equal to January 2, 1952 is feasible; but a date equal to January 32, 1952 is not. In the second case, the date of birth is erroneous. While this example is obvious, the "dubious" category calls for more explanation. Consider, for instance, a provider that has a graduation date of 1935. Assuming that he graduated from medical school at age 25, this makes him 84 years old in 1994. While there might be a few providers that are still practicing medicine at that age, it is unlikely. Therefore, the graduation date is classified as "dubious."

The five basic flag values are:

F – Feasible

M – Missing

E – Erroneous

D – Dubious

U – Unknown

Special flag values were created to meet specific circumstances, e.g., state license number is not right justified per *Registry* specifications. The criteria used to define feasible and non-feasible values for each variable are presented in Exhibit 5.

The flags for individual practice setting variables are created in the SAS program OUTINTS (see Panel B in Exhibit 3).[2] The flags for individual header variables are created in the SAS program OUTINTH. The names of the flag variables in the *Integrity* file have an "*I*" prefix. If there are both header and practice setting versions of a data element (e.g., date of birth), an "*H*" suffix indicates a header flag and a *S* suffix indicates a practice setting flag.

3.1.2 Logical Consistency

Logical consistency flags are based on the comparison of related variables, e.g., specialty code and credentials. If the specialty code indicates a podiatry specialty and the credentials indicates that the provider is a chiropractor, then there is a logical inconsistency between the two. There are four general values for the logical consistency flag variables:

- K – both variables are feasible and are logically consistent with each other
- I – both variables are feasible but are logically inconsistent
- M – both variables are missing
- N – any other combination

[2]Unless otherwise indicated, SAS programs referenced in Section 3.1 and 3.2 are listed in Panel B of Exhibit 3.

The one exception to the above list is for comparisons between graduation year and the open and close dates of medical schools. The reason is that closure dates are only available for medical schools and not for the other types of professional school. In those cases in which a provider graduated from a professional school other than a medical school, the flag value is set to a blank.

The flags for logical consistency comparisons between practice setting variables are created in the SAS program OUTLCFS. The flags for the logical consistency comparisons between header variables are created in the SAS program OUTLCFH. The names of logical consistency flag variables in the *Integrity* file have a "LCF" prefix.

3.1.3 Consistency of Header and Practice Setting Values

Consistency flags for the comparison of header and practice setting versions of a data element were created. The values of these variables should be identical. There are four general values for the header/practice setting consistency flag variables:

- K – both variables are feasible and have the same value
- I – both variables are feasible but have different values
- M – both variables are missing
- N – any other combination

The flags for the consistency between header and practice setting variables are created in the SAS program OUTHDSET. The name of the flag variables in the *Integrity* file have a "HDSET" prefix.

3.1.4 Consistency Across Practice Settings

Date of birth, date of death, school code, graduation year, credentials, physician status and the three sanction variables should have identical values over all practice settings. Consistency flags for the comparison of values across all practice settings were created. There are four general values for the across-practice setting consistency flag variables:

- K – all variables are feasible and have the same value
- I – all variables are feasible but at least one has a different value
- M – all variables are missing
- N – any other combination

In the event a provider has only one active practice setting, the value of the flag variable is set equal to a blank. The flags for the consistency of the comparison across practice settings are created in the SAS program OUTACROS. The names of the flag variables in the *Integrity* file have an "*A*" prefix.

3.2 Development of the Integrity File

The *Integrity* file contains both the unedited original variables and the integrity flags. It is developed in seven separate steps or programs. The programs and the files they create are described in Panel B in Exhibit 3. The first six programs create the integrity *flag* variables and resultant output files. The seventh program merges the six output files to create the *Integrity*

file. Because of the computer resource requirements for the computation performed, it was necessary to create six programs instead of one or two.[3]

The programs are all written in SAS and are located in a partitioned dataset maintained at the HCFA Data Center.[4] The seven programs output one file each to a cartridge and have the same dataset name (DSN) prefix. The names of the output files listed in Exhibit 3 serve as both the DSN suffix and the SAS internal name.

Only records (i.e., UPINs) with active practice settings are used in the six programs that create the flag variables. A variable, observation number (OBSNUM), was created to uniquely identify each active practice setting. This variable is necessary to properly link the "practice setting" output from the programs that create *practice setting* records. Unlike the other programs, programs 2, 4, and 6 (Panel B in Exhibit 3) produce one output record per provider. The name of the *Integrity* file is "SUPER" and it is sorted by UPIN, CARRIER, and MEDICARE (see Exhibit 4 for definitions). SAS variable labels were created for all critical variables.

3.3 Development of the Multiple-UPIN File

Only one SAS program is used in the development of the *Multiple-UPIN* file. The SAS program (OUTSCRNR) uses the unedited, original SAS dataset created by HER, as described in Section 2, and creates the SAS dataset SCREENER (see Panel C in Exhibit 3). The April 1994 version of SCREENER contains 156,017 active practice setting records representing 57,359 UPINs.

---

[3]Before SAS begins to process individual records, it reads the entire file into memory. Because the *Registry* file has over two million records, it requires an enormous amount of memory. To perform its computations and sorts, SAS needs to borrow large amounts of disk space to supplement memory. The result is that four programs took more than 30 minutes of CPU time. Users might want to consider using another programming language to create the *Integrity* and *Multiple-UPIN* files.

[4]See Exhibit 3 for names of datasets containing the programs. Contact HCFA to obtain copies of the programs (members) maintained in the datasets. The programs are also available on diskette from the National Technical Information Service.

The SAS program creates "key" variables to detect "suspect" multiple-UPIN cases. A key variable is created for each provider. The program then matches a provider's key variable to the key variables of other providers. If a match (i.e., the key values equal each other) is found, then the value of a 0,1 binary variable is set to one to indicate that there is another UPIN in the file with the same key value.

Four variants of the key variable are created and all are created in one run of the OUTSCRNR program. The names and components of each are shown in Exhibit 6. The corresponding names of the binary variables are also shown. In addition to the INKEYn binary variables (Exhibit 6), a special binary variable is created (INKEY123) in the same run of OUTSCRNR. A value of one indicates that any one of the first three key variables "detected" the "suspect" record. A value of zero indicates that it was only detected by the fourth key variable. The first key variable detects the fewest "suspect" cases while the fourth key variables detects the most – about eight times more than the first. The rationale for creating variants of the key variable can be found in Chapter 8 of the *Carrier Analysis*.

4.0 MAINTENANCE OF PROGRAMS USED TO DEVELOP THE INTEGRITY FILE

The criteria used to define feasible, erroneous, and dubious values for the date of birth, sanction year and graduation year need to be updated once a year (Exhibit 7). The first two should be updated when using the new April *Registry* file. The criteria for graduation year should be updated when using the new July *Registry* file, since graduation from medical school takes place in May and June. The year component of feasible values of date of death should be updated with the April file while the month component should be updated quarterly – see Exhibit 7 for details.

New flag variables can be created for other *Registry* variables (e.g., business zip code). If this is done, however, modifications then need to be made in the OUTSUPER program to modify the list of variables used to create the six count variables (COUNT_xx) which indicate

how many non-feasible flag values are present. The count variables, in turn, control the value of the binary variables used to determine if one of the integrity reports described in Section 5.1 should be printed. The programs listed in Exhibit 3 are internally documented to assist maintenance efforts.

5.0 REPORTS CREATED FROM THE INTEGRITY AND MULTIPLE-UPIN FILES

Two types of reports have been designed for use by HCFA, the Registry, and carriers for use with the *Integrity* and *Multiple-UPIN* files. From the *Integrity* file, several reports can be generated for each provider that has infeasible values in his/her *Registry* records. From the *Multiple-UPIN* file, reports can be generated for the suspect cases identified. Exhibit 8 shows the types of reports and the names of the SAS programs that create those reports. Each report is described in greater detail below. The reports are designed to be printed out on standard computer forms that have 132 columns and 66 lines.

5.1 Integrity File Reports

Reports are only produced for those providers that have infeasible or inconsistent values. The actual values of all variables are printed. The values of the flag variables, however, are printed only when the flag indicates a non-feasible, missing value or an inconsistent value. A blank, thus, is indicated in the event the flag variable is equal to F. On each report, just under the provider name, the number of non-feasible flag values is printed to indicate how many flag variables have non-feasible values.

5.1.1 Practice Setting Reports

There are two variants of the practice setting report – (1) the basic report and (2) the expanded report. The basic report has two blocks of variables – (1) the integrity of individual data elements and (2) the logical consistency flags for comparisons between variables. Both blocks of variables come from the practice setting section. The expanded report includes, in addition to the two variable blocks contained in the basic report, a block of variables indicating the consistency in coding between header and practice setting versions of variables. There is a provision in the programs for carriers to select only their providers, by using their carrier number. See the internal program documentation for details.

Unlike the other blocks of variables on the basic report, the flag values for the header variables are printed, because the basis of consistency definition is the equality of the data elements, rather than just both having feasible values. An example of the expanded report is provided in Exhibit 9. For Richard Doe, there is only one flag value to check – the header versus practice setting consistency flag for school code; it has a value of I. As can be seen on the exhibit, the school code on the practice setting indicates a foreign school while the header value indicates that the medical school is the Case Western.

In the practice settings reports, there are a few exceptions to the general rule that flag values are printed when they indicate non-feasible values. For the variable integrity flags, the "*M*" flag value is not printed for secondary specialty code or for primary and secondary board certification, because these three variables are not required by the *Registry*. Because secondary specialty is not required, the flag values for the logical comparisons between its value and those for credential, physician status and school code are not printed. The COUNT_xx variables in the OUTSUPER program do not count the missing values for secondary specialty and the board certification variables and the logical comparisons involving secondary specialty. As a result the PSFLAGx variables, used in the REPRTPSx programs to indicate that the practice setting has an integrity problem, do not reflect the occurrence of missing values for

these variables. The printing of the "M" flag value for these variables is suppressed in the REPRTPSx programs. The coding for the COUNT_xx variables in the OUTSUPER and the suppression of printing of "M" flag values in the REPRTPSx programs can be altered if needed.

5.1.2 Header Reports

There are two variants of the header report – (1) the basic report and (2) the expanded report. The basic report has two blocks of variables – (1) flags for individual data elements and their flag values and (2) logical consistency flags for comparisons between variables. These are similar to the first two blocks of variables on the practice settings reports, except the variables come from the header section of the record. The expanded report includes, in addition to the two variable blocks contained in the basic report, a block of variables testing for the equality of *practice setting* variables over all practice settings. The number of active practice settings is printed on the report, just above the last block of variables. The example of a header report in Exhibit 10 indicates that there is only one inconsistency. Specifically, the school codes on Dr. Doe's two practice settings differ.

5.2 Multiple-UPIN Report

An example of a *Multiple-UPIN* report is shown in Exhibit 11. It shows the values of variables that can assist in making a determination of whether further investigation is merited. The values printed for each UPIN come from its first active practice setting. In this example, the last name, first initial, school code, date of birth and PIN are identical for both UPINs. The only discrepancies are for the state license number and graduation year. One value is missing in each pair. It is likely that this is an actual case wherein one provider has two UPINs.

The key variable used to produce the example in Exhibit 11 was ID_KEY_1. The user of the program can use another key variable if so desired. Five lines of code in the SCRNRPT

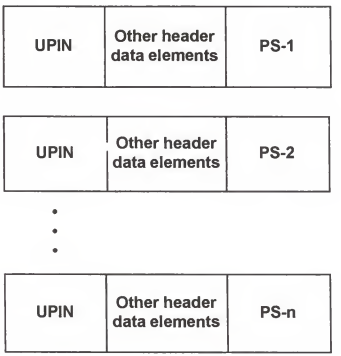
program need to be modified. ID_KEY_1 and INKEY1 need to be replaced by the appropriate variable pairs as indicated in Exhibit 6.

EXHIBIT 1

STRUCTURE OF A RECORD FOR AN INDIVIDUAL PROVIDER WITH MULTIPLE PRACTICE SETTINGS

Structure of a record on the file sent to HCFA by Transamerica:

Structure of records on HCFA files:



NOTE: PS denotes practice setting data elements (the suffix indicates the practice setting number).

upin\super\ex1.ppt\sjb

EXHIBIT 2

FORMATS OF REGISTRY DATA ELEMENTS

	CARRIER RECORD					*TOLIC FILE*	
Data Element	Required	COBOL Picture	Justification	Component Order	Missing Value	COBOL Picture	Component Order
Date of birth	Y	9(06)	R	MMDDYY	0	S9(09) COMP-3	YYMMDD
Date of death	Y	9(06)	R	MMDDYY	0	S9(09) COMP-3	YYMMDD
School code	Y	X(05)	R		blank	X(05)	
Year graduated	Y	9(02)	R	YY	0	S9(09) COMP-3	YY
Sanction code	N	X(01)			blank	X(01)	
Sanction, effective date	N	9(04)	R	MMYY	0	9(04)	MMYY
Sanction, number of years	N	9(02)	R		0	S9(03) COMP-3	
State license number	Y	X(12)	R		blank	X(12)	
Physician status	Y	9			0	X(01)	
Provider credentials	Y	X(03)	L		blank	X(03)	
Primary specialty	Y	9(02)	R		0	X(02)	
Secondary specialty	N	9(02)	R		0	X(02)	

SOURCES:

1. Carrier record formats from HCFA Transmittals.
2. Registry file formats from Transamerica Occidental Life Insurance Company: version of file sent to HCFA.

EXHIBIT 3

FLOW OF PROGRAMS TO CREATE THE INTEGRITY AND MULTIPLE-UPIN FILES

Program	Description	Output File
Panel A: Creation of the base SAS file from TOLIC's variable-length record file.		
1. --	BDMS creates a fixed-length record file.	MU00.@AAA2365.XUPIN.APR0494.FIX
2. SAS40494	Create a fixed-length record file in SAS format.	APR0494
Panel B: Creation of the *Integrity* file (first six programs use APR0494 as the input file).		
1. OUTINTS	Creates the integrity flag variables for individual practice setting data elements. It also writes out all of the original variables along with the new flag variables to a new file.	INTEGS
2. OUTINTH	Creates integrity flag variables for individual header data elements.	INTEGH
3. OUTLCFS	Creates the logical consistency flag variables for pairs of practice setting data elements (e.g., birth year and graduation year).	LCFS
4. OUTLCFH	Creates the logical consistency flag variables for pairs of header data elements (e.g., birth year and graduation year).	LCFH
5. OUTHDSET	Creates the consistency flag variables for comparisons between the header and practice setting versions of a data element (e.g., school code).	HEADSET
6. OUTACROS	Creates the consistency flag variables for comparisons of a data element (e.g., graduation year) across all of the provider's active practice settings.	ACROSS
7. OUTSUPER	This program merges together the 6 previously created files to create the *Integrity* file. It also creates variables that count the number of problems at several different levels.	SUPER
Panel C: Creation of the *Multiple-UPIN* file.		
1. OUTSCRNR	Create "key" variables to detect potential multiple-UPIN cases. Only UPINs with non-unique "keys" included in the SAS file. APR0494 is the input file for this job.	SCREENER

NOTES: 1. All program file are located in a partitioned dataset at HCFA: DSN = RR29.@DCH7910.HER.UPIN.LIB.

2. Except for the initial flat file created by BDMS, all SAS files are cataloged on cartridge. The output files all have the same *prefix* for the HCFA dataset name: DSN = RR00.@DCH7910.RR29.UPIN.APR0494. The internal SAS name is the same as the final qualifier (output file) in the dataset name.

3. Programs that utilize the *Integrity* and the *Multiple-UPIN* files are listed in Exhibit 8 and are briefly discussed in Section 5.

EXHIBIT 4

VARIABLE NAMES FOR REGISTRY DATA ELEMENTS

Variable	Description on Registry File	Code	HCFA Documentation Name (all names start with MPIER-)	SAS Name
Unique Physician Identification Number		H	UPIN-NUMBER	UPIN
Last name	Physician's last name	H	LAST-NAME	LNAME
First name	Physician's first name	H	FIRST-NAME	FNAME
Middle name	Physician's middle name	H	MIDDLE-NAME	MNAME
Name suffix	Physician's name suffix	H	GENERATION-SFX	SUFFIX
Date of birth	Date of birth (YYMMDD)	H	DATE-OF-BIRTH	DOB
Date of death	Date of death (YYMMDD)	H	DATE-OF-DEATH	DOD
Medical school code	XXX=state, YY=school	H	MED-SCHOOL-CODE	MEDSCHL
Year of graduation	Yr. physician graduated from med. school	H	YEAR-GRADUATED	GRADYR
UPIN sanction code	Sanction code this physician is under	H	UPIN-SANCTION-CODE	SANC1
Sanction effective date	Sanction effective dare (MMYY)	H	UPIN-SANCTION-EFFDATE	SANC1DAT
Years sanctioned	Number of yrs. physician sanctioned	H	UPIN-YEARS-SANCTIONED	SANC1YRS
Carrier number	The carrier number	S	CARRIER-NUMBER	CARRIER
Medicare billing number	Carrier-assigned PIN for this setting	S	PROVIDER-NO	MEDICARE
State license number	The state license number for this setting	S	ST-LICENSE-NUM	LICENSE
Physician's status	1=MD,DO 2=CH/DDM/DDS/DPM/OD	S	PHYS-STAT	PHYSSTAT
State in which physician is licensed		S	ST-LICENSED-IN	STLICENS
Physician's Credentials	Specify MD, DO, CH, DDM, DDS, DPM, or OD	S	PHYS-CREDENTIALS	CREDENTL
Primary specialty code	Specialty code from pt 3 of MCM	S	PRIM-SPEC	SPEC1
Primary board certification	Y or N or U for unknown	S	PRIM-BOARD-CRT	BOARD1
Secondary specialty code	Specialty code from pt 3 of MCM	S	SDRY-SPEC	SPEC2
Secondary board certification indicator	Y or N or U for unknown	S	SDRY-BOARD-CRT	BOARD2
Zip code-billing	Zip code of billing address	S	ZIP-BILLING	BILLZIP
Zip code-business	Zip code of business address	S	ZIP-BUSINESS	BIZZIP

EXHIBIT 4 (continued)

VARIABLE NAMES FOR REGISTRY DATA ELEMENTS

Variable	Description on Registry File	Code	HCFA Documentation Name (all names start with MPIER-)	SAS Name
Sanction code	Sanction code, if any, the physician is under	S	SANCTION-CODE	SANC2
Effective date of sanction	(MMYY)	S	SANCTION-EFFDATE	SANC2DAT
Years sanctioned	Length of sanction	S	YEARS-SANCTIONED	SANC2YRS
Resident or intern code	D= Deactivated R= Resident I= Intern P= Practic	S	RES-INTERN-CODE	INTERN
Group practice code	1=group 2=HMSA 3=both 4=Neither	S	GROUP-PRACT-CODE	PRACTIND
Physician's participation indicator	1=yes 2=no	S	PART-IND-1	PART1IND
Date the setting was added to Registry		S	SET-ADD-DATE	SETDATE
Last name for this setting	Physician's last name for this setting	S	SET-LAST-NAME	SETLNAME
First name for this setting	Physician's first name for this setting	S	SET-FIRST-NAME	SETFNAME
Middle name for this setting	Physician's middle name for this setting	S	SET-MIDDLE-NAME DATE 1	SETMNAME
Suffix name for this setting	Physician's suffix name for this setting	S	SET-GENERATION-SFX	SETSUFF
Date of birth for this setting	Physician's date of birth for this setting	S	SET-DATE-OF-BIRTH	SETDOB
Date of death for this setting	Physician's date of death for this setting	S	SET-DATE-OF-DEATH	SETDOD
Medical school code for this setting	Physician's medical school code for this setting	S	SET-MED-SCHOOL-CODE	SETMDSCH
Yr. physician graduated for this setting		S	SET-YEAR-GRADUATED	SETGRDYR

NOTES:

Code = H denotes header variable.

Code = S denotes practice setting variable.

EXHIBIT 5

DEFINITION OF DATA ELEMENT INTEGRITY MEASURES

Variable Definition	Variable(s)	Flag	Flag Definition	Criteria
Date of Birth (Setting & Header)	IDOBS IDOBH	F	Feasible	Year must be between 14 & 69 (inclusive), month must be feasible, and day must be feasible and consistent with month.
		D	Dubious	Year is between 01 & 13 (inclusive) or 99, month must be feasible, and day must be feasible and consistent with month.
		M	Missing	Value is equal to all 9's, 0 or blanks.
		1	Bad data	Value is equal to 10101 (January 1, 1901).
		E	Erroneous	Value is anything other than listed above.
Year of Birth (Setting & Header)	IDOBYRS IDOBYRH	F	Feasible	Year must be between 14 & 69 (inclusive).
		D	Dubious	Year is between 01 & 13 (inclusive) or 99.
		M	Missing	Value is equal to all 9's, 0 or blanks.
		E	Erroneous	Value is anything other than listed above.
Date of Death (Setting & Header)	IDODS IDODH	F	Feasible	Year is 93 or 94, month must be feasible, and day must be feasible and consistent with month.
		D	Dubious	Year is between 01 & 92 (inclusive), month must be feasible, and day must be feasible and consistent with month.
		9	Improper Coding	Date is 9 filled (999999).
		0	Alive	Date is 0.
		E	Erroneous	Value is anything other than listed above.
Year of Death (Setting & Header)	IDODYRS IDODYRH	F	Feasible	Year is either 93 or 94.
		D	Dubious	Year is between 01 and 92 (inclusive).
		9	Improper Coding	Date is 9 filled (999999).
		0	Alive	Date is 0.
		E	Erroneous	Value is anything other than listed above.
License Number	ILICENSE	F	Feasible	Value is feasible (ie, it is not any of the following).
		J	Justification	Value is not right justified.
		M	Missing	Value is blank-filled.
		Z	Zero's	Value is 0 filled.

EXHIBIT 5 (continued)

DEFINITION OF DATA ELEMENT INTEGRITY MEASURES

Variable Definition	Variable(s)	Flag	Flag Definition	Criteria
Medical School (Setting & Header)	IMEDSCHS IMEDSCHH	F	Feasible	Variable contains a feasible school code.
		J	Justification	Value is not properly justified.
		R	Retired	Variable contains a retired school code.
		M	Missing	Value is all blanks, all 0's, or 00001.
		3	Superceded Codes	Value is equal to 30000.
		4	Superceded Codes	Value is equal to 40000.
		5	Superceded Codes	Value is equal to 50000.
		E	Erroneous	Value is anything other than listed above.
Graduation Year (Setting & Header)	IGRADYRS IGRADYRH	F	Feasible	Year is between 40 & 93 (inclusive).
		D	Dubious	Year is between 30 & 39 (inclusive).
		M	Missing	Value is missing, 9's, or 0's.
		E	Erroneous	Value is anything other than listed above.
Primary Specialty Secondary Specialty	ISPEC1 ISPEC2	F	Feasible	Value is 01 to 08, 10 to 14, 18 to 20, 24 to 26, 28 to 30, 33 to 41, 16, 22, 44, 46, 48, 66, 70, 76 to 79, 81 to 86, 90 to 94 (all "to" series are inclusive).
		R	Retired	Variable contains retired osteopathic specialty code.
		M	Missing	Value is blanks, 0, or 00.
		U	Unknown	Value is equal to 99.
		S	Revised	Value is equal to 49.
		E	Erroneous	Value is anything other than listed above.
Sanction Code (Setting & Header)	ISANCS ISANCH	F	Feasible	Variable code is between A & R (inclusive).
		A	Non-event	Value is blank (UPIN not sanctioned).
		U	Unknown	Value is equal to U.
		E	Erroneous	Value is anything other than listed above.
Sanction Date (Setting & Header)	ISANCDTS ISANCDTH	F	Feasible	Year is between 66 & 94 (inclusive), and month is feasible.
		A	Non-event	Value is equal to blank or 0 (UPIN not sanctioned).
		E	Erroneous	Value is anything other than listed above.
Sanction Year (Setting & Header)	ISANCYYS ISANCYYH	F	Feasible	Year is between 66 & 94 (inclusive).
		A	Non-event	Value is equal to blank or 0 (UPIN not sanctioned).
		E	Erroneous	Value is anything other than listed above.

EXHIBIT 5 (continued)

DEFINITION OF DATA ELEMENT INTEGRITY MEASURES

Variable Definition	Variable(s)	Flag	Flag Definition	Criteria
Sanction Length (Setting & Header)	ISANCYRS ISANCYRH	F	Feasible	Value is between 1 & 15 (inclusive), or 20,25,30,35,40,45,50,99.
		A	Non-event	Value is equal to blank or 0 (UPIN not sanctioned).
		D	Dubious	Value is between 16 to 19, 21 to 24, 26 to 29, 31 to 34, 36 to 39, 41 to 44, 46 to 49, 51 to 98 (all "to" series are inclusive).
		E	Erroneous	Value is anything other than listed above.
Provider Credentials	ICREDENTL	F	Feasible	Value is equal to MD, DO, DDM, DDS, DPM, CH, or OD.
		M	Missing	Value is equal to blanks.
		E	Erroneous	Value is anything other than listed above.
Physician Status	IPHYSTAT	F	Feasible	Value is 1 or 2.
		M	Missing	Value is blank or 0.
		E	Erroneous	Value is anything other than listed above.
Primary Board Cert. Secondary Board Cert.	IBOARD1 IBOARD2	F	Feasible	Value is blank, Y, or N.
		U	Unknown	Unknown.
		E	Erroneous	Value is anything other than listed above.
Active Practice Indicator	IINTERN	F	Feasible	Value is R, I, or P.
		E	Erroneous	Value is anything other than listed above.
Group Practice Indicator	IPRCTIND	F	Feasible	Value is 1 or 4.
		M	Missing	Value is blank or 0.
		E	Erroneous	Value is anything other than listed above.
Participation Status	IPRT1IND	F	Feasible	Value is Y or N.
		M	Missing	Value is blank.
		E	Erroneous	Value is anything other than listed above.
State Licensed In	ISTLICNS	F	Feasible	Variable has a standard postal state abbreviation (including U.S. properties).
		M	Missing	Value is blank.
		E	Erroneous	Value is anything other than listed above.

EXHIBIT 6

COMPONENTS OF KEY VARIABLES FOR DETECTING MULTIPLE-UPIN CASES

Key Variable Name	Key Variable Components*	Matching Indicator
ID_KEY_1	DOB,L1_3,F1,LNAMESUM	INKEY1
ID_KEY_2	DOB,L1_3,LNAMESUM	INKEY2
ID_KEY_3	DOB,L1_3,F1	INKEY3
ID_KEY_4	DOB,L1_3	INKEY4

*Abbreviations:

DOB – Date of Birth
L1_3 – First 3 letters of last name
F1 – First initial of first name
LNAMESUM – Sum of the numbers assigned to the consonants of the last name (a value of one is assigned to B, two to C, three to D, and so forth, with 21 assigned to Z).

EXHIBIT 7

DATA ELEMENTS REQUIRING UPDATES OF BOUNDARY VALUES FOR FLAG VALUES: QUARTERLY FILE AND UPDATE FREQUENCY

Variable Flag Value	Quarter in which to change the Boundary Value		Update Frequency	Boundary Values of Year Components for April 1994 Integrity File		Increment Value
	Lower	Upper		Lower	Upper	
Birth Year						
Feasible	April	April	Annually	14	69	1
Dubious	Don't Change	April	Annually	(18)99	13	1
Date of Birth						
Feasible	April	April	Annually	14	69	1
Dubious	Don't Change	April	Annually	(18)99	13	1
Sanction Year						
Feasible	Don't Change	April	Annually	66	94	1
Sanction Date						
Feasible	Don't Change	April	Annually	66	94	1
Graduation Year						
Feasible	July	July	Annually	40	93	1
Dubious	July	July	Annually	30	39	1
Death Year						
Feasible	April	April	Annually	93	94	1
Dubious	Don't Change	April	Annually	01	92	1
Date of Death						
Feasible						
Year Component	April	April	Annually	93	94	1
Month Component*	Don't Change	All	Quarterly	01	3	3
Dubious						
Year Component	Don't Change	April	Annually	01	92	1
Month Component	Don't Change	Don't Change	N/A	N/A	N/A	N/A

N/A: Not Applicable.

*Upper boundary value of the month belonging to the upper year value needs to be incremented each quarter.

upin\super\EX7.XLS\ejb

EXHIBIT 8

INTEGRITY AND MULTIPLE-UPIN REPORTS

Report Name	SAS Program Name
Practice Setting	
Basic	REPRTPS1
Expanded	REPRTPS2
Header	
Basic	REPRTHD1
Expanded	REPRTHD2
Multiple-UPIN	SCRNRPT

NOTE: The practice setting and header reports use the *Integrity* file (see Panel B of Exhibit 3) as their input file. SCRNRPT uses the *Multiple-UPIN* file for its input data (see Panel C of Exhibit 3).

EXHIBIT 9
SAMPLE PRACTICE SETTING REPORT (EXPANDED)

THE SAS SYSTEM 14:24 TUESDAY, AUGUST 23, 1994
*** UPIN PRACTICE SETTING REPORT - EXPANDED ***

UPIN: KDD451 CARRIER: 64D PIN: 000DDDDDDXYF57

	LAST NAME	FIRST NAME	MIDDLE	SUFFIX
PRACTICE SETTING:	DOE	RICHARD	A	
REGISTRY (HEADER):	DOE	RICHARD	A	

NUMBER OF NON-FEASIBLE FLAG VALUES: 1

VARIABLE INTEGRITY

DATA ELEMENT	ACTUAL	FLAG	DATA ELEMENT	ACTUAL	FLAG
DATE OF BIRTH (YYMMDD)	410811		YEAR OF BIRTH	41	
DATE OF DEATH (YYMMDD)	0		YEAR OF DEATH	D	
SANCTION DATE (MMYY)	D		SCHOOL CODE	99999	
SANCTION YEAR	D		GRADUATION YEAR	67	
SANCTION LENGTH	0		PRIMARY SPECIALTY	10	
SANCTION CODE			SECONDARY SPECIALTY	DO	
GROUP PRACTICE	4		INTERN / RESIDENT	P	
CREDENTIAL	MD		STATE LICENSE NUMBER	DODODDDKZ476	
PHYSICIAN STATUS	1		STATE LICENSED IN	IA	
PRIMARY BOARD CERT. STATUS	U		SECONDARY BOARD CERT. STATUS	U	
PARTICIPATION STATUS	Y				

LOGICAL CONSISTENCY

DATA ELEMENTS	FLAG	DATA ELEMENTS	FLAG
PHYSICIAN STATUS & CREDENTIAL		BIRTH YEAR & GRADUATION YEAR	
SANCTION CODE & DATE		PROVIDER CREDENTIAL & PRIMARY SPECIALTY	
SANCTION CODE & LENGTH		PROVIDER CREDENTIAL & SECONDARY SPECIALTY	
SANCTION DATE & LENGTH		PROVIDER CREDENTIAL & SCHOOL CODE	
PHYSICIAN STATUS & SCHOOL CODE		SCHOOL OPEN/CLOSE YEAR & GRADUATION YEAR	
PHYSICIAN STATUS & PRIMARY SPECIALTY		PRIMARY SPECIALTY & SCHOOL CODE	
PHYSICIAN STATUS & SECONDARY SPECIALTY		SECONDARY SPECIALTY & SCHOOL CODE	

PRACTICE SETTING COMPARED TO HEADER VALUES

DATA ELEMENT	HEADER ACTUAL	HEADER FLAG	CONSISTENCY FLAG	DATA ELEMENT	HEADER ACTUAL	HEADER FLAG	CONSISTENCY FLAG
DATE OF BIRTH (YYMMDD)	41D811	F		YEAR OF BIRTH	41	F	
DATE OF DEATH (YYMMDD)	0	D		YEAR OF DEATH	0	D	
SCHOOL CODE	D38D6	F	I	GRADUATION YEAR	67	F	
SANCTION LENGTH	0	A		SANCTION DATE (MMYY)	D	A	
SANCTION CODE		A		SANCTION YEAR	D	A	

NOTE: The values of several variables were changed to preserve confidentiality.

EXHIBIT 10

SAMPLE HEADER REPORT (EXPANDED)

THE SAS SYSTEM 14:25 TUESDAY, AUGUST 23, 1994
*** UPIN HEADER REPORT - EXPANDED ***

UPIN: KO0451

LAST NAME	FIRST NAME	MIDDLE	SUFFIX
DOE	RICHARD	A	

NUMBER OF NON-FEASIBLE FLAG VALUES: 1

VARIABLE INTEGRITY	ACTUAL	FLAG
BIRTH YEAR	41	
DATE OF BIRTH (YYMMDD)	410811	
DEATH YEAR	0	
DATE OF DEATH (YYMMDD)	0	
SCHOOL CODE	03806	
GRADUATION YEAR	67	
SANCTION CODE		
SANCTION DATE (MMYY)	0	
SANCTION YEAR	0	
SANCTION LENGTH	0	

LOGICAL CONSISTENCY	FLAG
BIRTH YEAR & GRADUATION YEAR	
GRADUATION YEAR & SCHOOL OPEN/CLOSE DATE	
SANCTION CODE & DATE	
SANCTION CODE & LENGTH	
SANCTION DATE & LENGTH	

NUMBER OF ACTIVE PRACTICE SETTINGS: 2

CONSISTENCY ACROSS PRACTICE SETTING	FLAG
DATE OF BIRTH (YYMMDD)	
YEAR OF BIRTH	
DATE OF DEATH (YYMMDD)	
YEAR OF DEATH	
SCHOOL CODE	I
GRADUATION YEAR	
SANCTION CODE	
SANCTION DATE (MMYY)	
SANCTION YEAR	
SANCTION LENGTH	
CREDENTIALS	
PHYSICIAN STATUS	

NOTE: The values of several variables were changed to preserve confidentiality.

EXHIBIT 11

SAMPLE MULTIPLE-UPIN REPORT

*** POTENTIAL MULTIPLE UPIN CASES *** 12:14 TUESDAY, AUGUST 23, 1994 1

ID_KEY_1=030419DOEH42

VARIABLE NAME	DESCRIPTION	UPIN1	UPIN2	UPIN3
UPIN	Unique Physician Identification Number	J61599	J15591	
LNAME	Physicians last name	DOE	DOE	
FNAME	Physicians first name	HAROLD	H	
MNAME	Physicians middle name	I	I	
SUFFIX	Physicians name suffix			
DOB	Date of birth (YYMMDD)	030419	030419	
DOD	Date of death (YYMMDD)	0	0	
MEDSCHL	The school code	01875	01875	
GRADYR	Year phys. graduated from medical school	29	0	
CARRIER	The Carrier number	11260	11260	
MEDICARE	The Medicare billing number	000004567	000004567	
STLICENS	State in which the physician is licensed	MO	MO	
LICENSE	The state license number		0XY12	
PHYSSTAT	The physician status	1	1	
CREDENTL	The physicians credentials	DO	DO	
SPEC1	Physicians primary specialty	02	01	
BILLZIP	The billing zip code	68265282	68265000	
BIZZIP	The Business zip code	68265000	68265000	
INTERN	The resident or interm indicator	P	P	
SETDATE	Date the setting was added (YYYYMMDD)	19890419	19920618	
SETLNAME	Setting-Physicians last name	DOE	DOE	
SETFNAME	Setting-Physicians first name	HAROLD	H	
SETMNAME	Setting-Physicians middle name	I	I	
SETSUFF	Setting-Physicians suffix			
SETDOB	Setting-Physician date of birth (YYMMDD)	030419	030419	
SETDOD	Setting-Physician date of death (YYMMDD)	0	0	
SETMDSCH	Setting-The school code	01875	01875	
SETGRDYR	Setting-Year phys grad from med school	29	0	

NOTE: The values of several data element were changed to preserve confidentiality. ID_KEY_1 was used to identify this case.

CMS LIBRARY
3 8095 00014112 3